Development is in the details

Citation for published version:
Mottus, R & Rozgonjuk, D 2019, 'Development is in the details: Investigating age differences in the Big Five domains, facets and nuances', *Journal of Personality and Social Psychology*, vol. N/A, pp. 1-14. https://doi.org/10.1037/pspp0000276

Digital Object Identifier (DOI):
10.1037/pspp0000276

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Personality and Social Psychology

Accepted to *Journal of Personality and Social Psychology* (24/10/2019)

Development is in the details: Age differences in the Big Five domains, facets and nuances

René Mõttus [1,2,*], Dmitri Rozgonjuk [2,3]

[1] Department of Psychology, University of Edinburgh, Edinburgh, UK

[2] Institute of Psychology, University of Tartu, Tartu, Estonia

[3] Center of IT Impact Studies, Johann Skytte Institute of Political Studies

Abstract

We examined the extent to which the Big Five domains, 30 facets and nuances (uniquely represented by individual questionnaire items) capture age differences in personality, expecting domains to contain the least and nuances the most age-related information. We used an Internet sample (N = 24,000), evenly distributed between ages of 18 and 50 years and tested with a 300-item questionnaire. Separately based on domains, facets and items, we trained models to predict age in one part of the sample and tested their predictive accuracy in another part. Big Five domains predicted age with an accuracy of r = .28, whereas facets' (r = .44) and items' (r = .65) predictions were more accurate. Less than 15% of the sample was needed to train models to their optimal accuracy. Residualizing the 300 items for all facets had no impact on their predictive accuracy, suggesting that age differences in specific behaviors, thoughts and feelings (i.e., items) were not due to domains and facets but mostly unique to nuances. These findings replicated in a multi-sample dataset tested with another questionnaire. We found little evidence that age differences only appeared nuanced because items referred to age-graded roles or experiences. Therefore, a substantial part of personality development may be uniquely ascribed to narrow personality characteristics, suggesting the possibility for a many-dimensional representation of personality development. Besides theoretical implications, we provide concrete illustrations of how this can open new research avenues by enabling to study systematic variations between traits.

Keywords: Personality development; age differences; nuances; facets; machine learning

[*] Corresponding author: René Mõttus, Department of Psychology, University of Edinburgh, Edinburgh, UK; rene.mottus@ed.ac.uk

Authors' Note: Authors are grateful to Lew Goldberg for creating IPIP, to John A. Johnson for collecting the data and making them publicly available, to Jonah Sinick for calculating re-test correlations and to Taavi Unt for methodological advice.

Development is in the details: Age differences in the Big Five domains, facets and nuances

One of the main goals of personality science is to understand how individual differences in behavior, thinking, feeling and motivation change as individuals grow older – how personality develops. This is most typically studied by describing age differences in the domains of the Five-Factor Model (FFM) or the Big Five (e.g., Roberts, Walton, & Viechtbauer, 2006) and, with longitudinal data available, their rank-order stability (e.g., Roberts & DelVecchio, 2000). Additionally, individual-level changes in these traits are sometimes linked with other individual-level variables such as life experiences or social roles (e.g., Bleidorn, Hopwood, & Lucas, 2018). On average, older adults appear to have socially more mature personality trait levels than younger adults (Caspi, Roberts, & Shiner, 2005), individual differences in the traits tend to be stable over several years (e.g., Terracciano, Costa, & McCrae, 2006) and specific correlates of trait changes often remain elusive (Bleidorn et al., 2018; Denissen, Luhman, Chung & Bleidorn, 2019). Here, we take a fresh and unprecedentedly detailed look at age differences in personality trait scores from young to middle adulthood: between ages 18 and 50 years.

Age differences in the Big Five domains and facets

In terms of the Big Five domains, middle-aged adults are often found to score lower in Neuroticism and higher in Agreeableness and Conscientiousness than young adults (e.g., McCrae, Costa, de Lima et al., 1999; Soto, John, Gosling, & Potter, 2011), in both cross-sectional and longitudinal studies (e.g., Terracciano, McCrae, Brant, & Costa, 2005) and across a range of cultures (e.g., McCrae, Terracciano et al., 2005; Donnellan & Lucas, 2008; Lucas & Donnellan, 2009). The trends are often not linear from young to middle adulthood, although their exact shapes vary considerably from study to study. Mean age differences are less robust for Extraversion and Openness, with many studies reporting negative overall age trends for these traits in adulthood (e.g., McCrae et al., 1999; Donnellan & Lucas, 2008; Lucas & Donnellan, 2009) but some reporting the opposite or no systematic age differences at all (e.g., Soto et al., 2011; Ashton & Lee, 2016).

But there is more to personality development than can be described with the Big Five domains. Personality traits form a hierarchy, where the Big Five domains are only one level and split into narrower traits such as aspects or facets (DeYoung, 2015; Soto & John, 2017; McCrae & Mõttus, in press). These narrower traits reveal additional information regarding age differences in personality.

Specifically, facet-level analyses often show that facets of the same FFM domains have different developmental trajectories. In a meta-analysis by Roberts and colleagues (2006), the Social Vitality facet of Extraversion declined slightly over adulthood, whereas the Social Dominance facet had an upward trend. Terracciano and colleagues (2005) found cross-sectional decrease trends in the Competence and Achievement Striving facets of Conscientiousness but an increase trend for the Dutifulness facet; in their longitudinal analyses, mean scores of the Deliberation facet increased throughout the life-course, whereas the Competence and Self-Discipline facets increased until middle age and then declined. Jackson, Walton, Harms and colleagues (2009) also reported systematic variability in age differences across Conscientiousness facets. Soto and John (2012) reported longitudinal decreases for the Rumination and Depression facets of Neuroticism but not for the Anxiety and Irritability facets; in their cross-sectional analyses, only Rumination scores were lower in older as opposed to younger adults. Likewise, there is substantial variability in the magnitudes of age differences among the facets of Openness and Agreeableness (e.g., Terracciano et al., 2005; Lucas & Donnellan, 2019; Bleidorn, Kandler, Riemann, Spinath, & Angleitner, 2009; Soto & John, 2012; Ashton & Lee, 2016). According to Soto and colleagues (2011, p. 342) a "… growing body of findings indicates that conceptualizing traits at the level of Big Five facets is necessary for a full understanding of life span age differences in personality; research at the domain level can provide a rough sketch of these differences, but not a complete picture".

Age differences may extend to nuances

But facets may not be the lowest level of the personality trait hierarchy. There appear to be even narrower traits, called nuances, that can be indexed by single personality questionnaire items, at least as far as our current ability to conceptualize the below-the-facets personality variance goes (McCrae, 2015; McCrae & Mõttus, in press; Mõttus, Kandler, Bleidorn, Riemann, & McCrae, 2017). Hundreds of nuances (i.e., operationally, items) appear to have properties of traits such as stability over time, agreement across assessment methods, heritability and predictive validity (Mõttus et al., 2018; Seeboth & Mõttus, 2018). This is not because the nuances are merely markers of broader traits such as the Big Five domains and facets because the trait properties also apply to their *unique* variance – that is, when the variance of the Big Five domains and facets has been partialled out. McCrae (2015) has estimated that two-thirds of the valid variance in a typical personality test item is unique and only one third represents the broader trait the item is designed to measure. If so, nuances abound.

What is important here is that nuances may also have distinct developmental trajectories. For example, Mõttus and colleagues (2015) found in a sample of Estonian adults that none of the 30 facets of the NEO Personality Inventory (NEO-PI; McCrae & Costa, 2010) met the criteria for measurement invariance across adulthood because individual items had developmental trajectories that could not be accounted for by the Big Five domains and facets. Also, combining samples from several countries and languages, Mõttus and colleagues (2018) found that the unique variance in over a third of the 240 NEO-PI items showed significant correlations with age and that these correlations tended to at least moderately replicate across demographically diverse country-samples.

These findings are preliminary but they suggest that personality development may be more nuanced than how it is usually represented. However, there is currently little, if any at all, research that systematically compares different levels of the personality trait hierarchy (e.g., domains, facets and nuances) in their degrees of capturing age differences in personality. This is what the current study sets out to explore. Do age differences mostly pertain to the Big Five domains (i.e., variance shared by items designed to measure each), with unique variance in facets contributing some extra details? If so, and the unique variance in nuances only captures a small additional fraction of age differences, there would be little value in studying personality development at the nuances level. If, however, a substantial proportion of age differences turns out to be unique to nuances and not properly mapped by either domains or facets, this could have important implications for the study of personality development, both in terms of describing and explaining it.

Potential implications of age differences in nuances

First, ubiquitous age differences in nuances would suggest that for a "complete picture" (Soto et al., 2011; p. 342), personality changes ought to be described at the level of narrower traits than the Big Five, as is typically done, or even at the level of their facets (Mõttus et al., 2015). It could simply be unwise to gloss over details when these can tell a rich story of psychological development, even if this story is more complicated than we are accustomed to thinking. Instinctively, many researchers may be repelled by specificity – all else equal, parsimony is a good principle. But it would not be "all else equal", if much of personality development indeed happened at the nuance-level. It is important to realize that a nuanced description of personality changes would not invalidate findings at the domain or facet level but only extend and enrich them. Nuances allow, but do not mandate, "zooming in".

Second, this could explain why studies into age differences in some personality domains and facets vary in findings (Costa, McCrae, & Löckenhoff, 2019). For example, while Extraversion tends to decline throughout adulthood when measured with the NEO-PI (McCrae et al., 1999), the trait shows virtually no mean-level differences when described with the Big Five Inventory (BFI; Soto et al., 2011) and may even trend upwards when measured with the HEXACO Personality Inventory (Ashton & Lee, 2016); discrepant findings were also found for Openness in these studies. Varying age trajectories have also been reported for similarly labeled but differently measured facets such as liveliness/activity (Terracciano et al., 2005; Lucas & Donnellan, 2009; Soto et al., 2011; Ashton & Lee, 2016) or anxiety (Terracciano et al., 2005; Soto et al., 2011). This could be because the instruments sample different nuances of the traits.

Third, if age differences are to a substantial extent specific to nuances, this could suggest that the mechanisms of personality development, and thereby psychological development more generally, are heterogeneous and to a commensurate degree narrowly-acting – that is, specific to narrow traits (Soto & John, 2012). On the one hand, this could inform us of the limits of general explanations for personality development such as an overarching social "maturation" (Caspi et al., 2005). On the other hand, developmental mechanisms could be general even with nuance-specific age trajectories but not aligned with how the nuances coalesce into domains and facets in cross-sectional individual differences. For example, the nuances of the same domains and even facets can socially mature (a general explanation) but do so at different rates. Either way, a substantial degree of nuance-specificity of developmental trajectories can explain why it has been difficult to pinpoint genetic variants responsible for trait variance (e.g., Lo et al., 2017) or experiential correlates of trait change (e.g., Bleidorn et al., 2018): the associations have almost exclusively been studied using broad aggregate traits, whereas narrowly-acting effects may cancel out in them.

Fourth, although a substantial degree of nuance-specificity in developmental mechanisms may seem like bad news at first glance, we argue that this may unlock entirely new ways of studying personality development. Specifically, we may need to let go of the standard practice of focusing on how individual traits vary with age and examine their patterns instead. For one, mapping nuances that change at similar rates or correlate in their changes (when longitudinal data is available) and studying their shared properties (e.g., a common psychological domain such as affect or motivation) can reveal developmental mechanisms. Moreover, correlating rates of change in a large and diverse set of traits with potential explanatory variables that have also been quantified for these traits would allow setting up and testing hypotheses that are hard to even imagine when describing personality change with just a few traits.

For a simple example, this would allow us to numerically test the hypothesis that personality development reflects social maturation (Caspi et al., 2005): we could quantify, say, 100 diverse nuances in mean-level differences on the one hand and degrees of reflecting social maturity on the other (e.g., using expert ratings or correlations with objective maturity-criteria), and then expect these two properties to correlate across the 100 nuances. This would be a powerful alternative to eyeballing mean-level change patterns in traits such as the Big Five domains and judging that these *look* like people are generally becoming socially more mature (Caspi et al., 2005) – we could numerically estimate the extent to which this is the case. Also, Wilt and Revelle (2015) have shown that traits can be quantified in the degrees to which they represent affect, behavior, cognition and motivation: doing this for a diverse set of nuances and correlating the results with age-trajectories in these nuances can show whether personality development pertains equally to these psychological domains or whether personality varies with age more in, say, the motivational than the behavioral domain. Among other things, these questions can be separately addressed in various age levels and the resultant findings compared: perhaps the maturation trend only pertains to specific life periods or psychological domains change differently in various stages of development.

For yet another example, we may hypothesize that personality development in adolescents reflects catching up with social expectations (social pressure) for behavior, thinking, feeling and motivation (Denissen, Van Aken, Penke, & Wood, 2013). For this, we could characterize a diverse set of nuances in their mean-level trajectories and/or changes in variance as well as quantify social expectations for them (e.g., from parents, teachers, peers) and expect that a) mean-level changes in the traits are proportional to social expectations for them and b) traits with stronger expectations decrease comparatively more in variance over time as adolescents are increasingly pressured to be alike.

Addressing these and other hypotheses based on mapping out systematic differences between traits can be a major boost for personality theories, but they inevitably require a many-dimensional (e.g., nuances-based) model of personality change. Instead of mapping change trajectories and their possible explanatory mechanisms for just, say, five traits, we will need a larger *sample* of traits. For a parallel, we cannot study individual differences with a sample of just five people.

Current study

To compare the degrees to which age differences in personality can be ascribed to different levels of the personality hierarchy, we relied on a large internet sample tested using a comprehensive 300-item personality questionnaire designed to measure the Big Five domains and their 30 facets (IPIP-NEO; Goldberg, 1999). Specifically, we "predicted" age from either the Big Five domains, their 30 facets or the 300 items selected to index these traits, and correlated the predicted ages with actual ages. Based on previous findings with the NEO-PI (e.g., Mõttus et al., 2015, 2018), we hypothesized that many of the 300 items would be markers of nuances, besides being markers to their intended facets and domains, and therefore capture unique developmental information. To directly quantify the degree to which it was the unique variance in items (uniquely reflecting nuances) that correlated with age rather than the variance they shared with domains and facets, we compared the age-predictive power of items to that of their residuals, after being regressed on all Big Five domains and facets.

Predicting age, as a dependent variable, from personality traits as independent variables may seem an unexpected choice at first glance – should causality not run the other way around? However, it is important to realize that a prediction model is not necessarily a causal model but merely a useful statistical tool for summarizing how much information one set of variables (personality traits) collectively contains about another variable (age), accounting for the overlaps among the former. If facets collectively capture more variance in age than domains, they are likely to have developmental trajectories distinct from those of the domains they belong to; if nuances (uniquely represented by items) collectively capture more variance in age than facets, they likely have developmental trajectories distinct from those of the facets they belong to. Besides, the extent to which residualizing items for facets (and thereby automatically for domains) reduces their collective ability to predict age is a direct measure of how much of the age-relevant information in them is due to the facets and domains they are designed to measure.

Another way to think of this procedure is that we constructed bespoke "polytrait" scores specifically tailored to represent age differences in personality, based on either Big Five domains, facets, the 300 items or their residuals, and then tested how strongly these actually aligned with age (Mõttus et al., 2018; Vainik et al., in press). The polytrait scores for age were the weighted sum of all variables included in the scores, with the weights being the degrees to which these variables were linked with age. In domains-based polytrait scores, facets and nuances of the same domains were assumed to have similar associations with age; in facet-based polytrait scores, nuances of the same facets were assumed to have similar associations with age, but nuances of different facets even within the same domain could vary in their age-associations. In polytrait scores directly based on items or their residuals, the nuances uniquely represented by these items were allowed to have different associations with age regardless of whether they belonged to the same broader trait or not. This way, it is easy to see how our procedure tested the overall extent to which nuances within the same domains or facets had similar or unique age trajectories – or put differently, the overall dimensionality of age-differences. This is similar to how geneticists routinely create polygenic scores for various phenotypes from different sets of single nucleotide polymorphisms and then compare them in how much variance in the target phenotypes they collectively capture (Plomin & Stumm, 2018).

To account for model over-fitting and more complex models *a priori* capturing more age-related information, we "trained" the models (i.e., created the weights for polytrait scores) and applied them (i.e., tested polytrait scores-actual age correlations) in independent subsamples. For model training, we employed elastic net regularized regression (Zou & Hastie, 2005) that focuses on the unique contribution of each predictor and strives for parsimony by shrinking smaller coefficients (i.e., trait-age associations) to zero, which also helps against over-fitting. In a simulation, Seeboth and Mõttus (2018) showed that this procedure is not biased against broader traits: if the associations of personality with another variable (here, age) pertain to domains/facets as opposed to nuances, then the polytrait scores based on domains/facets indeed correlate with the variable more strongly than the polytrait scores based on nuances. For comparison and robustness check, we also created polytrait scores based on zero-order correlations of age with domains, facets, items and their residuals, not accounting for overlap among the personality variables and not penalizing small associations.

We replicated the main findings using a combination of different datasets and a different personality test, NEO-PI. Specifically, we created polytrait scores for age based on items and item residuals using item-age and item residual-age correlations from a published multi-country meta-analysis and correlated the polytrait scores with age in an independent sample. This allowed us to estimate the proportion of age-related unique variance in items free of sampling and, to a large degree, language overlaps between the model training and validation phases, in addition to testing the robustness of the findings across personality tests. This was a very rigorous test of whether age-differences in personality are higher-dimensional than can be represented with the Big Five domains and their facets.

METHOD

We only used previously (anonymously) collected and publicly available datasets (see Johnson, 2014, https://osf.io/tbmh5; Goldberg, 2018, https://doi.org/10.7910/DVN/HE6LJR); therefore, we did not apply for ethics approval for neither the main nor the replication analyses. The list of publications using the data employed in the main analyses is provided in the Open Science Framework (https://osf.io/4zydh); none of them used the data for a purpose in any way similar to ours. To our knowledge, the data used for the replication analyses has never been used for a similar purpose either.

Main analyses

Participants

From among the 307,313 participants in the initial sample (185,149 women; mean age 25.19 years, *SD* = 10.00), describing themselves as belonging to 238 countries, we selected those aged between 18 and 50 years and from mostly English-speaking countries (USA, Canada, UK, Australia, Ireland and New Zealand; to reduce possible biases due to participants not speaking English as their native language). Also, as participants at younger ages were over-represented in the sample and this could bias age differences estimates towards trends in younger ages, we stratified the sample so that there would be an equal number of men and women in six age groups (18-25, 26-30, 31-35, 36-40, 41-45, and 46-50 years). We strove to select 2,000 men and 2,000 women into each age group (hence the target sample was *N* = 24,000) by randomly sampling from among participants with complete data and targeted genders and age ranges. However, in the the two oldest age groups of men only 1,674 and 1,225, respectively, had complete data; we therefore sampled participants into these groups from among those with up to two missing responses and median-replaced their missing responses.

Measures

Participants completed a 300-item personality questionnaire (IPIP-NEO; Goldberg, 1999) that is based on International Personality Item Pool (Goldberg, 1999; Goldberg, Johnson, Eber et al., 2006) and mimics the structure of the NEO-PI in measuring the Big Five domains and their 30 facets; each facet was measured with 10 items instead of eight as in NEO-PI (McCrae & Costa, 2010). A 120-item subset of these 300 items is often used as a shorter but structurally similar questionnaire, containing four items for each of the 30 facets instead of 10 (Johnson, 2014); the 120 items were selected by eliminating those containing more unique variance beyond the 30 facets, hence they are collectively expected to contribute less towards the measurement of nuances than the 300 items. In addition to the full set of 300 items, we separately considered the degree to which this subset of 120 items contained age-related information. This was to explore whether the additional 180 items conferred incremental age-descriptive value, suggesting even higher dimensionality in personality-age associations than can be captured by the 120 items alone. In addition, information about the respondents' age, gender and self-reported country was provided.

Analysis

Data analyses were conducted in R statistical environment 3.6.1 (R Core Team, 2019). We first created five sets of predictors of age and transformed them to *z*-scores: the Big Five domains, their 30 facets, the 300 items of the full IPIP-NEO, the residuals of the 300 items after being regressed on the 30 facets (the item being residualized was excluded from its facet at the time; Mõttus et al., 2017, 2018), and the 120 items of the shorter version of the test. Then, using a series of (linear) elastic net regressions as implemented in the *glmnet* package

(Friedman, Hastie, Simon, & Tibshirani, 2016), we predicted age from these five sets of personality variables and calculated Spearman correlations of the predicted ages with the observed ages. We relied on a shrinkage (also know as regularized or penalized) regression approach such as elastic net as opposed to an ordinary least squares/maximum likelihood regression because it is specifically designed for large numbers of inter-correlated predictors. It shrinks regression coefficients towards zero – many exactly to zero – so that effectively only the strongest unique predictors are retained in the model. As a result, it efficiently deals with overlaps among predictors, including their multicollinearity, and yields more parsimonious prediction models that tend to replicate better because of the natural tendency of regression models to produce inflated coefficients due to model overfitting (Yarkoni & Westfall, 2017).

Elastic net is related to the ridge (Hoerl & Kennard, 1970) and Least Absolute Shrinkage and Selection Operator (LASSO) regressions (Tibshirani, 2011). The former applies a shrinkage penalty to regression coefficients that depends on the sum of the squares of these coefficients; it tends not to increase model parsimony as it rarely shrinks regression coefficients to zero. LASSO regression applies a penalty that depends on the sum of the absolute values of these coefficients and leads to more parsimonious models. LASSO regression has the downside of often randomly selecting one of many correlated predictors and setting coefficients for others to zero, which means that LASSO solutions are not unique, whereas ridge regression tends to shrink coefficients for correlated predictors towards each other (Waldmann et al., 2013; Friedman, Hastie & Tibshirani, 2010). Elastic net (Zou & Hastie, 2005) combines the ridge and LASSO penalties, mitigating limitations associated with each of them alone (e.g., Chapman et al., 2016). It yields parsimonious models (due to the LASSO penalty), in which groups of strongly correlated predictors are treated in the same way (due to ridge penalty), all being either included or excluded from the model (Waldmann et al., 2013; Zou & Hastie, 2005). The optimal regularization parameter lambda (λ) was obtained using 10-fold cross-validation within the training sample such that it minimized cross-validated error across the folds.

Importantly, not all of the 24,000 participants were used for training such models; instead, we trained the models in smaller participant subsets and applied them to predict actual age in the remaining participants. We varied the training sample size from 1% to 75% of the 24,000 participants, with 1% increments. This is because we were interested in the optimal sample sizes for obtaining stable (elastic net) models with up to hundreds of predictors, as the results could guide future studies. At each training sample size, models were calculated 100 times with random sample permutations. When correlating the predicted and actual ages, we therefore obtained 500 correlations for each training sample size: 100 for domains, 100 for facets, 100 for 120 items, 100 for 300 items and 100 for the residuals of the 300 items.

Finally, we attempted to unravel the "black box" of the elastic net procedure with a less sophisticated and highly tractable approach. First, in order to gauge the incremental predictive value provided by elastic net regression coefficients over the simplest possible models based on zero-order correlations, we calculated correlations of domains, facets, 300 items and their residuals with age in a training subsample (67% of the total sample) and applied these to predict age in the remaining participants (by multiplying correlations with the standardized scores of the respective variables and summing the products). The extent to which associations obtained this way fell short of the elastic net-based predictions showed the added value of applying the shrinkage regression algorithm. Second, we correlated the predicted ages based on the elastic net regression to those based on zero-order correlations of items and their residuals with age. We expected the elastic net to yield predictions of age that were closer to those based on item residuals rather than to those based on raw item scores because both of these ways of estimating item-age associations disregarded variance shared among items. If so, this would show how the elastic net approach worked and how it provided its superior predictive power (should this happen).

Of note is that we tested for measurement invariance of the domains and facets across the six age groups used for sampling participants, using the guidelines of Chen (2007) and treating items as indicators in factor models for facets and facets' sum-scores as indicators in factor models of domains (using maximum likelihood estimator, scaling latent traits by setting their variance at unity and not specifying residual correlations among indicators). All domains and facets met the criteria for weak (factor loading equality across age groups), strong (intercept equality) and strict (residual variance equality) invariance according to the Standardized Root Mean Square Residual (SRMR) criterion (Chen, 2007). According to the Comparative Fit Index (CFI) criterion of Chen (2007), Extraversion, Openness and Agreeableness domains (ΔCFI = .014 to .054) and 16 facets (ΔCFI = .012 to .027; most notably E5: Excitement-Seeking, C4: Achievement-Striving and O2: Artistic interests, with ΔCFI > .02) failed the strong invariance test; also three of the facets failed the strict invariance test. However, these violations of measurement invariance were much smaller than has been reported for NEO-PI (Mõttus et al., 2015); lack of strong invariance is consistent with face- and item-specific age-trends. That we could establish weak invariance suggests that the facet and domain scores were qualitatively comparable across disparage age groups.

Replication analyses

To replicate the key finding – the proportion of age-relevant information that would be in items' unique variance – across samples, languages and measurement instrument, we used the correlations of age with the 240 items of the NEO-PI and their residuals (controlling for all facets and FFM domains, as in the main analyses) to predict age in independent people. Among other properties of the NEO-PI items and their residuals, their age-correlations were reported by Mõttus and colleagues (2018); these were meta-analytic correlations that had been calculated across five samples from Canada, the Czech Republic, Denmark, Germany and the United States tested in five different languages (total N = 5,615, mean ages ranging from 34 to 57 years).

We used these meta-analytic correlations to compare the age-predictive power of items and their residuals in an independent sample only tested in English: the Eugene Springfield Community Sample (ESC; N = 857; Goldberg, 2018). After replacing up to 40 missing responses per participant with medians, we could use data from 840 ESC participants. With a mean age of 50.63 years (SD = 13.08; 56% women) the ESC participants were on average somewhat older than the participants in the meta-analytic sample. Specifically, we multiplied the meta-analytic age-correlations with the standardized NEO-PI item scores and their residuals in the ESC and correlated the resulting polytrait scores with ESCS participants' actual ages. Training and validating models in different and demographically diverse samples provided a particularly stringent test of the age-predictive power of items' unique variance (i.e., nuances).

Availability of data, analytic code and materials

Data and the R code for the main and replication analyses are available at the Open Science Framework (https://osf.io/4zydh), making the empirical work fully reproducible. The personality test used in the main analyses (IPIP-NEO) is available in the public domain (https://ipip.ori.org/newNEOFacetsKey.htm), whereas the test used in the replication analyses (NEO-PI) is proprietary and cannot be made available; rephrased items are provided in Mõttus and colleagues (2018; Online Supplemental Material).

RESULTS

Main analyses

Spearman correlations between raw item scores and their residuals varied from .56 to .94 with a mean of .77. For an alternative metric, this means that about 59% of items' variance was unique, on average, including measurement error. The associations of individual domains, facets, items and their residuals with age are given in the Online Supplemental Material as a) correlations and b) linear bivariate model-implied differences between ages 18 and 50 years in standard score units. Items' correlations with facets and domains are also reported in the Online Supplemental Material, but note that the items were not removed from their facets for these correlations.

Elastic net results

The results of the elastic net-based analyses comparing the Big Five domains, facets, items and item residuals in predicting age in independent participants are depicted in Figure 1 (across different training sample sizes) and Table 1 (for a single sample size). Four main results are worth noting.

Table 1. *Correlations between actual ages and ages predicted by elastic net models based on the Big Five domains, facets, items and their residuals.*

	Big Five	Facets	120 items	300 items	Residuals of 300 items
Mean	.28	.44	.54	.65	.65
SD	< .01	< .01	< .01	< .01	< .01

NOTE: The mean and standard deviation (*SD*) are across 100 replications with the training sample of 67% of the total sample.

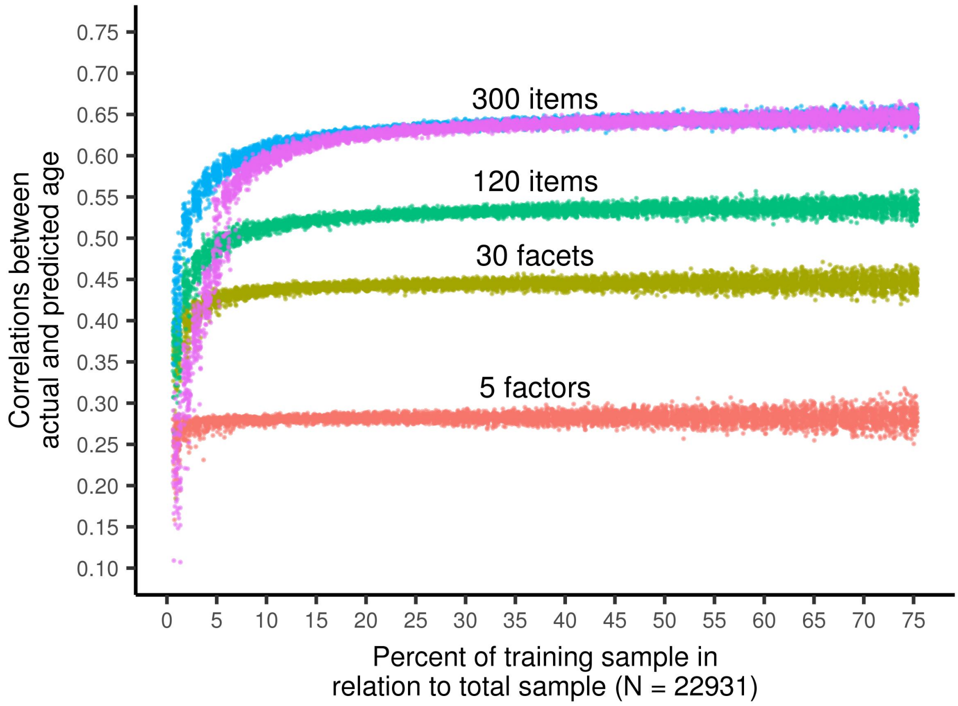

Figure 1. *Correlations of actual and predicted ages based on different levels of the personality trait hierarchy (domains, facets, and nuances as indexed by individual questionnaire items) and across different training sample sizes. For 300 items, the purple (darker) dots are for the correlations based on their residuals, whereas the blue dots are for raw item scores.*

First, the Big Five domain scores excluded over half of the age-sensitive information available in the data. While the accuracy of the Big Five-based predictions averaged at $r = .28$, the 30 facet-based predictions had an average of $r = .44$, and the 300 items based predictions achieved more than two times higher correlations with age, with average $r = .65$. These are fairly dramatic increments for both facets and nuances over domains (and nuances over facets), especially for a questionnaire that had been carefully designed to measure only the Big Five domains and facets in the first place.

Second, the principle of more information allowing for better prediction extended beyond the comparison of different levels of the personality hierarchy (domains, facets and nuances as operationalized by individual items) because models based on 300 items invariably out-predicted those based on 120 items. This shows that there may be many tens of nuances that are uniquely associated with age, above and beyond domains, facets and even the nuances already captured by the 120 items in the shorter IPIP-NEO version. Again, these results are especially notable because the instrument had been (implicitly) designed so that such findings would *not* emerge: items of the same IPIP-NEO facets have been to a large extent selected to be redundant and interchangeable (i.e., to measure the same attribute and yield high internal consistency) and thereby be similarly linked with external variables.

Third, although models with more variables needed larger training samples to yield stable estimates (i.e., to have low variance among the predicted age-observed age correlations for similar models; Figure 1), the main pattern was almost invariably observable with no more than 500 participants (about 2% of our sample), except for models based on item residuals which needed slightly more participants. With 1,500 participants in the training sample, most estimates had almost stabilized; only models based on item residuals needed around 3,000 participants to stabilize. Of course, the fewer predictors, the faster stable predictions could be achieved – complex models have an inevitable cost. But the cost should be tolerable, especially given that studies into personality development increasingly boast sample sizes in thousands.

But it is the fourth finding that may be the most dramatic. Residualizing the 300 items for all 30 facets scores had virtually no impact on their collective ability to capture age-sensitive information. That is, regression models trained based on items that no longer contained the variance of the Big Five domains and facets allowed for just as strong out-sample prediction of age as models based on raw item scores that did contain the domain and facet variance. This suggests that the associations of items with age could generally not have been due to the Big Five domains and facets. In other words, individuals at different ages differed in specific behavioral, cognitive, affective and motivational patterns indexed by individual items not because they differed in the domains and facets *per se* but because they differed in something that the items uniquely reflected – that is, nuances.

This result also mitigates the concerns that findings pointing to the nuanced-ness of age differences resulted from more complex models *a priori* out-predicting simpler ones. This is because models based on items and their residuals had equal numbers of predictors.

Nuanced associations

A Manhattan plot (Figure 2) where the correlations of the 300 items and their residuals with age are grouped by the Big Five domains and facets illustrates why item-based models captured more age-related information than those based on domains and facets. Items of the same facets, never mind the same domains, often varied notably in their correlations with age, with significant item-age correlations varying in direction for 15 of the 30 facets. Although individual correlations with age were generally of smaller magnitude for item residuals than for raw item scores, significant associations abound, with 152 of the 300 item residual-age correlations significant at $p < .05$ after Holm correction for multiple testing (Holm, 1979).

The absolute correlations for raw item scores ranged up to .25 with a mean of .07; for item residuals, the maximum and mean were .16 and .03, respectively. To represent these effects in the time metrics (see Online Supplemental Material), on average, 18- and 50-year-olds differed by 0.25 standard deviations in raw item scores (maximum 0.86) and 0.10 standard deviations in item residuals (maximum 0.52). For comparison, age differences between the 18- and 50-year-olds averaged 0.34 standard deviations for facets (maximum 0.94) and 0.39 standard deviations for the Big Five domains (maximum 0.71). Such a pattern – individual effect sizes generally being smaller for more specific traits (especially for nuances represented by items and their residuals) combined with the vastly superior amount of information that they *collectively* captured about age – is consistent with unique age differences being distributed across a large number of traits.

We do not intend to describe the nuance-level associations in great detail, because our main aim is to simply show that they are numerous, which in itself may tell us even more about psychological development than pinpointing any one association. However, we do highlight some correlations that defied the trend among their associated items (see Online Supplemental Material for all associations). For example, although nine items of the A2: Morality facet had positive correlations with age, older respondents tended to score lower on an item referring to not cheating in taxes; perhaps concern for institutions declines with age, while that for individuals increases. Items of the A5: Modesty facet that referred to dislike of being in the center of attention and boasting decreased with age, but older people were somewhat more likely to believe to know answers to their questions, perhaps referring to increases in self-perceived wisdom. The C2: Orderliness facet items (referring to organized surroundings and daily life) generally trended upwards with age, but the one referring to liking things to be "just right" had a downward trend; perhaps increasing age means striking, on average, a better balance between order and perfection. Although the C4: Achievement Striving items that referred to working hard and with full effort often correlated positively with age, older participants reported a lower urge to succeed; perhaps reasons for working hard become increasingly intrinsic with age. This replicates a previous observation regarding age differences in the items of NEO-PI Achievement Striving facet (Mõttus et al., 2015). The E1: Friendliness facet contained items with both small positive and negative correlations with age, with older participants reporting it harder to make friends, to warm up quickly to others, and to cheer up people but being more interested in and feeling more comfortable around others. Also, E3: Assertiveness contained items with age trajectories in different directions, with older participants more likely to take charge in situations but less likely to try to influence others and to talk them into doing things. The items of the E4: Activity Level that referred to reacting quickly bucked the trend for older people generally being more (pro-)active. Several items of E6: Cheerfulness had negative correlations with age (older people reported having less fun and laugh less), whereas the item "look at the bright side of life" had a positive association with age; perhaps older people are better in emotion regulation, despite (feeling the need for) less joy.

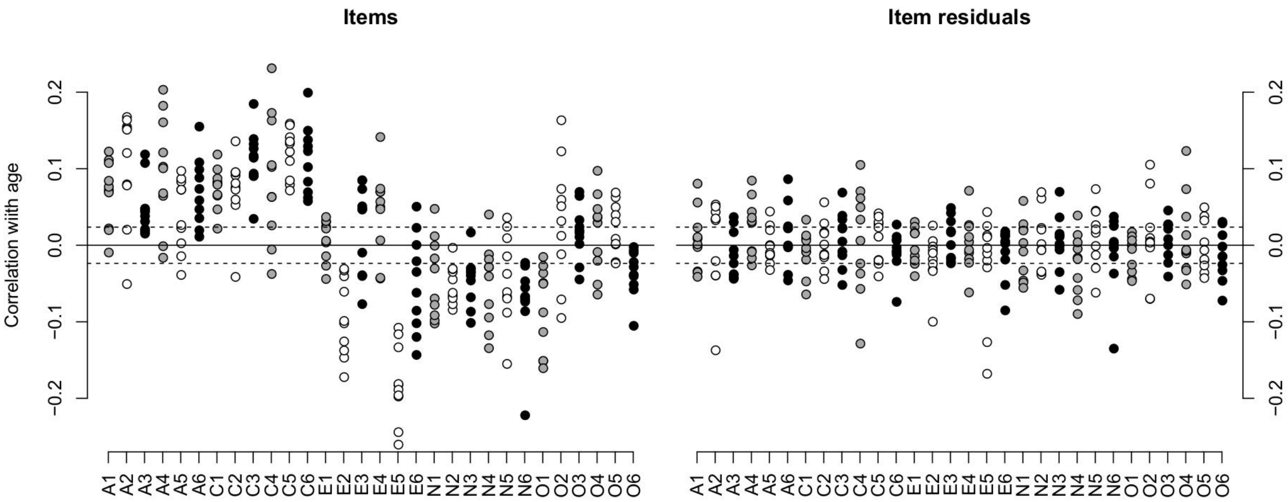

Figure 2. *Manhattan plots for the correlations of the 300 items and their residuals with age. The correlations are grouped according to the Big Five domains (indicated by letter) and their facets (indicated by number; see ipip.ori.org/newNEOFacetsKey.htm for facet names). The dashed line indicates the threshold for statistical significance after Holm correction for multiple testing (Holm, 1979). The plots were drawn using the* psych *package (Revelle, 2018).*

As for N1: Anxiety, the scores of several items correlated negatively with age (less fear and worry), whereas older participants tended to report being less relaxed; perhaps coming of age means less of the intense negative emotions, although not necessarily less and maybe even more general stress. The scores of several items of the N4: Self-Consciousness and N5: Immoderation facets trended downwards with age, but older participants reported being more bothered with difficult social situations as well as eating and spending too much. Two O2: Artistic Interests facet items (liking music and concerts) bucked the trend for positive correlations in items referring to liking art, dance, nature and museums. The O3: Emotionality facet was a mix of positive and negative correlations, with older age being linked with noticing and understanding emotions but having and enjoying fewer of them. For O4: Adventurousness, an interesting pattern emerged, whereby older participants liked the *idea* of change but still reported being attached to conventional ways and not liking to visit new places; perhaps the types of changes people are after tend to vary with age. For O5: Intellect, age was positively correlated with items referring to liking mental challenges and being able to handle a lot of information but not with items referring to abstract ideas and philosophical discussions; perhaps older age means cognitive engagement with more concrete, down-to-earth problems rather than flights of philosophical fantasy.

As for the strongest associations, several were from the E5: Excitement-Seeking facet, which was among the most internally consistent scales in how its items correlated with age. For example, higher age meant especially strong dislike for loud music, bungee jumping, crowds and excitement. Besides these, items with the strongest upward age trends referred to being able to make up one's mind, doing more work than needed to get by, not insulting people or getting back at them, not doing crazy things, avoiding misrepresenting facts, and liking flowers but not surprise parties.

In combination, these nuanced associations point to a possibility that younger age is generally associated with intense, swift action-packed expressions of personality *across* broader trait domains, whereas increasing age is associated with growing moderation, realism and deliberately negotiating with social and non-social demands. If so, then what changes with age may not be particular domains *per se* but the ways of relating to life that permeates them.

Were unique age differences in items measurement artifacts?

It is possible that the age differences uniquely captured by items were in part measurement artifacts: items may refer to age-graded social roles, life-style aspects and other circumstances, in addition to representing personality *per se*. Sports, wild parties and loud music may be more relevant for young people than middle-aged adults, for example. We addressed this possibility in three ways.

First, we checked the content of the items whose residuals had the strongest correlations with age, complementing the overview of the distinct item-age correlations outlined in the previous section. Of the top 30 items with the strongest residual correlations with age, four could refer to age-circumscribed behaviors at facet value: liking bungee jumping, surprise parties or loud music, and cheating on taxes that possibly goes with substantial income and is therefor more likely in older age. Also, it is possible that several items of the O2: Artistic Interests facets with the strongest residual age-correlations could be considered age-circumscribed: liking musing, flowers, poetry and the beauty of nature. However, a majority of the top 30 items did not refer to age-specific roles or circumstances: being able to make up one's mind, being motivated to succeed, preferring variety to routine, valuing cooperation over competition, or being afraid of mistakes, among others.

Second, we re-ran the elastic net models, excluding the items of three facets that contributed many items with residual-age correlations and could have the most age-circumscribed content: E5: Excitement-Seeking, O2: Artistic Interests and C4: Achievement Striving (100 random permutations with 67% of the sample for training). On average, domains, facets, the 300 items and their residuals predicted age with r = .26, .40, .61, and .61, respectively. These results suggest that it was not merely the unique variance in the possibly most age-circumscribed items that underpinned the general trend for items to contain unique age-related information.

Third, we compared the predictive ability of domains, facets, items and their residuals in four narrower age groups: ages 18 to 20 years, 21 to 30 years, 31 to 40 years and 41 to 50 years (100 random permutations with 67% of the sample for training in each case). This is because social roles, life-style aspects and other circumstances are less likely to systematically differ between immediately adjacent ages than across more distant ages: for example, although listening to loud music or doing crazy things may be less socially normative for middle-aged adults than for 18-year-olds, these are likely equally pertinent to those aged 18, 19 and 20 years. Therefore, items retaining their age-predictive advantage within narrow age groups would suggest that this is not merely due to items capturing extra-personality content. Indeed, in all four age groups items and their residuals outperformed the Big Five domains and facets (Table 2). For example, items had on average 40% to 67% stronger correlations with age than facets, which is similar to or even higher than the results obtained across the whole age range (Table 1). Of course, the overall magnitude of age differences was smaller in narrower age groups because of less variance in age and it decreased from 20s to 30s to 40s, suggesting non-linearity in age-trends.

In combination, these analyses suggest that the age-relevant information that items contained beyond domains and facets was not merely reflective of age differences in social roles, life-style aspects and other age-specific circumstances.

Table 2. *Correlations between actual age and age predicted by elastic net models based on the Big Five domains, facets, items and their residuals in four age groups.*

	Big Five	Facets	300 items	Residuals of 300 items
		Ages 18 to 20 years		
Mean	0.08	0.15	0.21	0.18
SD	0.02	0.02	0.01	0.01
		Ages 21 to 30 years		
Mean	0.10	0.21	0.35	0.34
SD	0.01	0.01	0.01	0.01
		Ages 31 to 40 years		
Mean	0.11	0.16	0.24	0.21
SD	0.01	0.01	0.01	0.02
		Ages 41 to 50 years		
Mean	0.07	0.12	0.19	0.17
SD	0.01	0.01	0.01	0.01

NOTE: The mean and standard deviation (*SD*) are across 100 replications with the training sample of 67% of the total sample.

Analyses with zero-order correlations

We correlated age with the Big Five domains and their facets as well as with the 120 and 300 items and the residuals of the 300 items in 67% of the total sample (N = 24,000 * .67 = 16,080) and then applied these correlations to create corresponding polytrait scores in the remainder of the participants. To obtain the polytrait scores based on each set of predictors, standardized scores of each predictor were multiplied by its correlation with age in the training sample and the resulting products were summed. The correlations of the polytrait predictions and actual ages are given in Table 3. A pattern emerged that was similar to the findings based on the elastic net modes, with 300 items allowing for a stronger prediction of age than the domains, facets and 120 items.

However, there were two notable differences. First, the correlation-based associations were generally substantially weaker than the elastic net-based associations (Table 1), suggesting that elastic net models did provide a better way of representing trait-outcome associations. Second, the item residuals-based correlations provided substantially stronger correlations with actual age than predictions based on raw item scores. This may seem surprising in suggesting that the facet and domain variance in items had constrained their associations with age (not in absolute value as can be seen in Figure 2 but in terms of the collective informativeness of the associations). But this may also explain why elastic net models clearly out-performed correlation-based models: these models took item co-variance (i.e., variance underpinning domains and facets) into account.

Table 3. *Correlations between actual age and age predicted from it correlations with the Big Five domains, facets, items and their residuals.*

	Big Five	Facets	120 items	300 items	Residuals of 300 items
Mean	.26	.29	.34	.36	.49
SD	< .01	< .01	< .01	.01	< .01

NOTE: The mean and standard deviation (*SD*) are across 100 replications with the training sample of 67% of the total sample.

Understanding the models and their implications for personality-age associations

Item residuals-based correlations allowing for better prediction of age in independent participants than raw item scores-based correlations suggest that the former could be more in line with how age was predicted from the elastic net models. Indeed, the elastic net-based predictions of age correlated with item residuals-based predictions at $r = .74$ but with item raw scores-based predictions at $r = .55$ (these analyses were based on 300 items and with 67% training sample). However, because elastic net did not residualize variables but only estimated regression coefficients by removing shared variance and then applied these to raw item scores for prediction, we mimicked the same procedure by creating polytrait scores for age such that correlations obtained from residuals in the training sample were applied to raw item scores in the remainder of participants. Indeed, the resultant polytrait scores correlated highly with elastic net-based predictions ($r = .94$). They also correlated substantially with actual age ($r = .61$), which was similar to the elastic net-based prediction of age ($r = .65$).

This suggests that the elastic net models yielded the best prediction of age by effectively removing much of the impact of the Big Five domain and facet variance from regression coefficients and explains why elastic net models based on item residuals performed fairly similarly to models based on raw item scores. In simple terms, the domain and facet variance was not needed in items for the prediction of age and could, in fact, even hinder outlining their most informative associations.

Reliability of items

In Mõttus and colleagues (2018), retest reliability of items was estimated using the myPersonality dataset (Kosinski et al., 2015), in which a subset of participants had completed a 100-item FFM measure (Goldberg at al., 2006) multiple times; incidentally, 62 of the 100 items were those of the IPIP-NEO used in the current study. We calculated correlations for each of these 62 items across three retesting intervals: 2 to 7 days, 8 to 14 days, 15 to 21 days. The average retest correlations were $r = .68$, .67, and .66, respectively (the individual estimates are given in Online Supplemental Material; $N = 970$ to 1,639). These estimates are comparable, if slightly higher, to those of the NEO-PI items referred to by McCrae and Mõttus (in press; footnote 5). We therefore conclude that the reliability of a typical questionnaire item is in .60s. This is of course less than the typical retest reliability of domains and facets (over .80; McCrae and Mõttus, in press) but likely higher than many researchers would suspect (Wood, Nye, & Saucier., 2010). There was inconsistent evidence for items with stronger retest reliability to have stronger absolute correlations with age (Spearman $\rho = .14$ to .27, $p = .037$ to .280, for the three retest intervals; $N_{\text{items}} = 62$).

Replication analyses

Could the key finding that much of the age-related information was in items' unique variance be specific to the IPIP-NEO or biased due to a consistent testing language and sampling homogeneity (the training and validation subsamples were drawn from the same larger sample)? To find out, we used the meta-analytic correlations of age with the 240 items of the NEO-PI and their residuals (controlling for all facets; Mõttus et al., 2018) to predict the actual ages of the ESCS participants.

Polytrait scores based on applying item-age correlations to standardized items tracked with the ESCS participants' ages at $r = .33$, whereas applying item residuals-based age-correlations to standardized item residuals resulted in polytrait scores that correlated with observed ages at $r = .43$; these were slightly lower than the corresponding estimates in Table 3. However, applying item residuals-based correlations to standardized items yielded the highest accuracy, with $r = .50$.

Therefore, the correlations were somewhat weaker in the NEO-PI-based replication than in the IPIP-NEO-based original findings, which can be explained with fewer items and the meta-analytic correlations being based on a combination of different samples that did not entirely match the age range of ESC. But the pattern of findings was clearly replicated: the unique variance in items was mostly responsible for their collective correlations with age rather than the variance associated with the FFM domains and facets.

DISCUSSION

The glass of our results may be seen as half full, half empty. On the one hand, the scores of the Big Five domains could account for a substantial degree of age differences in a diverse set of specific personality traits as operationalized with hundreds of test items. Three hundred items allowed the prediction of age with the accuracy of about .65 and five domains with the accuracy of about .28, suggesting that the domains could do about 43% of what the items did. Given the simplicity of domain-based models, this may be impressive. But it is also not surprising, because the 300 items had been carefully selected to measure these domains in the first place and the domain scores consisted of nothing but these very items.

On the other hand, the domains did leave more than half of the age differences in items unaccounted for. In other words, the majority of information on how behavior, thinking, feeling and motivation vary with age is discarded when only operationalizing personality with the Big Five. There clearly is a lot more to personality development than changes in the Big Five, even in the information collected for the measurement of the Big Five in the first place. Researchers may choose to pursue this additional information or be content with what the domains can do. Either way, we now have a better understanding of what these choices entail in terms of capturing age differences.

Relying on the set of 30 facet-level traits instead of the Big Five domains closed the gap between how much age-related variance was in domains and items by about a half. This is a substantial improvement over domains and suggests that rigorous studies into personality development ought to measure personality with at least a few dozens of facet-like traits (Terracciano et al., 2005; Soto et al., 2011). However, even facets may not be sufficient for an entirely comprehensive account of personality development because items collectively contained a substantial amount of additional age-sensitive information. This suggests that it may take several dozens, perhaps even hundreds of trait constructs to fully capture developmental trajectories in behavior, thinking, feeling and motivation. The finding that 300 items designed to measure the Big Five domains and facets contained substantially more age-relevant information than the subset of (presumably the best) 120 items selected to measure exactly the same set of traits provides compelling evidence for this.

Descriptive and explanatory implications

Of course, what constitutes a "complete picture" (Soto et al., 2011, p. 342) of personality trait development is an open question: does it have to address every minute behavioral change or is a broad-stroke description of general change patterns good enough a sketch? We propose that there is no single answer because different research purposes beget accounts of development with different levels of comprehensiveness and detail. This is a classic bandwidth-fidelity dilemma, in which details have to be traded against parsimony.

For many purposes, the picture provided by the Big Five domains is likely just good enough. For instance, engaging the public in personality research is helped by simple personality representations and many properties of personality such as age differences in the rank-order stability of traits (Wagner, Lüdke, & Robitzch, 2019) can be equally well described using any trait model. For other purposes, we may want to harness as much information about personality development as possible. We prefer not to appear prescriptive as to which personality hierarchy level to rely on when describing personality development, but we emphasize that there do exist various legitimate choices – perhaps more choices than has been realized before, in that researchers can legitimately move beyond facets and study how personality develops at the nuances level. However instrumental, the Big Five and even their facets are not necessarily the only nor always the best options for representing changes in personality.

Besides excluding a substantial degree of age-related variance, the Big Five domains and facets *per se* may not explain much of the age-differences in personality. This is because residualizing items for the Big Five domains and facets had virtually no impact on their degrees of capturing age-related variance. Collectively, items allowed for just as strong prediction of age in independent people when their age-associations had been calculated free of the variance of the Big Five domains and facets compared to when the associations contained the variance of these higher-order traits. This puzzling finding may indicate that much of the development in behavior, thinking, feeling and motivation is, in fact, driven by mechanisms unrelated to the Big Five and their facets – plausibly by something uniquely captured by nuances. If so, besides less than a complete description of change, the Big Five domains and facets may be of somewhat limited help for explaining personality development. This echoes a similar comment by Baumert and colleagues (2017). This finding is also similar to that of Seeboth and Mõttus (2018) pertaining to how items were correlated with a range of life outcomes: residualizing the Big Five items for the Big Five scores had limited impact on their predictive value.

New research avenues provided by many-dimensional models of personality development

That much of personality development is in the detail may seem like a bad news at first glance because there appears to be an overwhelming number of traits and possible mechanisms to consider. However, we prefer to see this as an opportunity, because many-dimensional (nuances-based) representations of personality development may open entirely new research avenues.

Besides harnessing more of systematic developmental variance between individuals, a many-dimensional model of development would allow for studying systematic *variations between traits* in how they develop and intersect with different types of developmental influences and outcomes. This would provide researchers with an opportunity to think of and test hypotheses that are impossible to consider with a small number of traits. In the introduction, for example, we provided illustrations in relation to the personality maturation hypothesis (Caspi et al., 2005), the idea that mean-level development may reflect gradual compliance with social expectations (Denissen et al., 2013) or whether age differences pertain most strongly to affective, cognitive, behavioral or motivational aspects of personality (cf. Wilt & Revelle, 2015).

Also, the selection of item-level correlations with age presented above provides basis for a hypothesis that younger age is associated with more intense and action-packed expressions of personality, whereas increasing age can be characterized by more moderation and deliberate balancing of social and non-social demands – across many broad trait domains. This hypothesis can be tested by characterizing nuances (e.g., items) in their degrees of (behavioral) intensity and/or requiring deliberate consideration of social/non-social demands and then correlating these degrees with the age-trajectories of the items. For such strategies to be most fruitful, researchers would benefit from personality measures that sample as broad a range of nuances as possible (McCrae & Mõttus, in press).

Do items really capture age differences in personality traits?

We suspect some readers may think that nuances do not really represent personality but perhaps stylistic variations or (characteristic) adaptations of more "core" personality traits (McAdams & Pals, 2006; DeYoung, 2015; McCrae & Sutin, 2018). We sympathize with this interpretation but think that this is almost a value judgment. If we define traits as individual differences characteristics that are stable over time, observable across assessment methods, predictive of life outcomes and perhaps in part heritable (e.g., Funder, 1991; McCrae &

Costa, 2008), then nuances are just as much traits as the Big Five domains or facets (McCrae, 2015; McCrae & Mõttus, in press; Mõttus et al., 2017, 2018). The current findings add distinct developmental trajectories to the list of the trait-properties of nuances.

Another interpretation of age differences being the strongest in items could be that items capture age difference in social roles, life-style aspects or other circumstances, on top of the traits they are designed to be indicators of. Essentially, then, age differences in items' unique variance could be partly or fully measurement artifacts. We addressed this possibility in three ways and found limited support for it. First, at face value, majority of the items displaying comparatively strongest unique associations with age did not refer to anything age-circumscribed, although some did. Second, excluding the items of the facets that could contain the most age-specific content (Excitement-Seeking, Achievement Striving and Artistic Interests) did not change the main pattern at all – items' (unique) variance collectively still captured more information about age than the Big Five domains and facets. Third, this pattern was also present when analyses were constrained to narrow age ranges such as 18 to 20 years or 21 to 30 years: it is unlikely that social roles, life-styles and other circumstances change as dramatically within a few years as they could change over decades.

But imagine we did accept that the unique variance in items represented mostly (characteristic) adaptations or measurement artifacts rather than traits *per se*. If so, the current findings would be even more surprising. In this case, age differences in personality could be almost entirely ascribed to age differences in the adaptations or measurement artifacts, whereas traits as such would vary very little with age. Personality would then be set like a plaster already by 18 years of age! If so, some theories of personality trait development would need a major overhaul (e.g., Roberts & Nickel, 2017). Fortunately, we do not have to accept this.

Measurement of nuances: Going forward

The findings suggest that personality varies with age along more numerous and narrower dimensions than can be represented by one of the most comprehensive systems of facets (Costa & McCrae, 1992). But the fact is that we currently do not have any bespoke model of these narrow dimensions, nuances. Even the very evidence for their existence is based on the "scraps" of the Big Five domain and facet measures – the leftover variance in individual items specifically designed *not* to measure nuances. Clearly, we will need an empirically based and comprehensive taxonomy of nuances and tools for measuring them if we are to tap into the currently almost hidden but vast personality variance, be it for predicting outcomes (Seeboth & Mõttus, 2018), studying personality development, or other purposes.

One way forward would be to restart the kinds of taxonomic research programs that lead to the Big Five, either based on the lexical approach or starting from unstructured item pools but not with the goal to identify the few *major* but the *many* dimensions of personality (Wood et al., 2010). Historically, the goal has been parsimony, achieved by aggregating and filtering out as much information as possible (Booth & Murray, 2018). For the development of a taxonomy of nuances, however, the goal should be capturing as much of individual differences in personality traits as is possible with measurement tools that are still usable.

Among other possibilities, we think that this can be achieved by creating measurement instruments that explicitly focus on the psychometric quality of single items such as their unambiguity and construct-relevance, retest reliability, cross-rater agreement and low evaluativeness, while also avoiding wasteful redundancy among items. For example, the retest reliability of many items is in the .70s (Online Supplemental Material; McCrae & Mõttus, in press), whereas it is considerably lower for many others. Reasonably high reliability is thus achievable for even single items, but this has rarely been the goal so far. Explicitly relying on only high-quality items will allow measuring more traits with fewer items. We imagine that it would be possible to measure at least 100 nuances with, say, 200 or 300 good-quality items. Naturally, these nuances could still be aggregated into broader traits such as the facets and domains of the Big Five or HEXACO, thereby fully recognizing the hierarchical structure of personality traits and simultaneously allowing for both parsimonious and comprehensive representations of personality variance. In fact, a systematic and comprehensive pool of nuances could also improve the measurement of facets and domains because their breadth and comprehensiveness would improve. This could thus be a win-win strategy.

Limitations

We point to five limitations of the study, although there may undoubtedly be many more. First, we only modeled linear associations with age, although there are almost always non-linear trends as well. This means that our models likely underestimated the overall degree to which personality varies as a function of age.[1] However, linear associations were sufficient for our main purpose – comparing the degree to which different levels of the personality trait hierarchy capture age-relevant information. Second, the 300 items were unlikely to exhaust the pool of possible personality nuances or even provide a representative sample of them. After all, the items had been carefully selected *not* to measure anything but the Big Five domains and facets. This probably means that there is yet more age-sensitive information above and beyond the domains and facets than we could show. Using personality measures that are not *a priori* structured to measure a set of specific traits such as California Child Q-Set (Block & Block, 1980) or Inventory of Individual Differences in the Lexicon (Wood et al., 2010) may be helpful, but ultimately, measures based on bespoke and comprehensive models of nuances will ultimately be required.

Third, our main analyses relied on an Internet sample which was unlikely to be representative of the general population and may have even been differentially unrepresentative at different ages. However, although this could have distorted the overall associations with age, this was probably less of a problem for comparing the domains, facets and nuances because the participants were always the same. Likewise, the main findings held across diverse samples for which data had been collected in more conventional ways (replication analyses). Fourth, we only studied cross-sectional age differences, which may have confounded cohort differences with personality change. It is possible that cohort differences are especially pronounced in narrow traits. Future studies should directly measure longitudinal personality change in nuances, preferably over multiple time-points. For example, this could be achieved by controlling for the variance of domains and facets in nuances (items or multi-item nuance scales) and mapping out the changes in them. Finally, we only considered ages between 18 and 50 years, whereas a substantial amount of personality changes is bound to take place prior to and following this period of life. For example, it may be that the incremental value of nuances is smaller in childhood or old age as developmental processes may be more coupled then.

Conclusion

To conclude, the question is not whether there is personality beyond the Big Five and their facets: yes, there *is* a lot more of it by any formal criteria. Instead, the question is whether and, perhaps even more accurately, when and how to address it in such a way that it enriches our description and understanding of personality and its development. We do not think that many-dimensional models of personality and its development can or should be useful for each and every purpose. For describing major patterns, the Big Five may be an excellent fit. But we provided robust evidence that personality varies with age along many more dimensions and gave concrete examples of how capitalizing on this many-dimensionality of age differences can help us tackle novel research questions.

[1] In the Online Supplemental Material, we present age differences separately for ages 18 to 34 years and 34 to 50 years; this allows estimating the non-linearity in age trajectories in the Big Five domains, facets, items and their residuals.

REFERENCES

Ashton, M. C., & Lee, K. (2016). Age trends in HEXACO-PI-R self-reports. *Journal of Research in Personality*, *64*, 102–111. https://doi.org/10.1016/j.jrp.2016.08.008

Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P., … Wrzus, C. (2017). Integrating Personality Structure, Personality Process, and Personality Development. *European Journal of Personality*, *31*, 503–528. https://doi.org/10.1002/per.2115

Bleidorn, W., Hopwood, C. J., & Lucas, R. E. (2018). Life Events and Personality Trait Change. *Journal of Personality*, *86*, 83–96. https://doi.org/10.1111/jopy.12286

Bleidorn, W., Kandler, C., Riemann, R., Spinath, F. M., & Angleitner, A. (2009). Patterns and sources of adult personality development: growth curve analyses of the NEO PI-R scales in a longitudinal twin study. *Journal of Personality and Social Psychology*, *97*, 142–155. https://doi.org/10.1037/a0015434

Block, J., & Block, J. H. (1980). *The California Child Q-Set.* Palo Alto, CA: Consulting Psychologists Press.

Booth, T., & Murray, A. L. (2018). How Factor Analysis Has Shaped Personality Trait Psychology. In *The Wiley Handbook of Psychometric Testing* (pp. 933–951). https://doi.org/10.1002/9781118489772.ch29

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality Development: Stability and Change. *Annual Review of Psychology*, *56*, 453–484. https://doi.org/10.1146/annurev.psych.55.090902.141913

Chapman, B. P., Weiss, A., & Duberstein, P. R. (2016). Statistical learning theory for high dimensional prediction: Application to criterion-keyed scale development. *Psychological Methods, 21,* 603-620. https://doi.org/10.1037/met0000088

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling, 14,* 464–504. https://doi.org/10.1080/10705510701301834

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, Fl.: Psychological Assessment Resources.

Costa, P. T., McCrae, R. R., & Löckenhoff, C. E. (2019). Personality Across the Life Span. *Annual Review of Psychology*, *70*, 423–448. https://doi.org/10.1146/annurev-psych-010418-103244

Denissen, J. J. A., Van Aken, M. A. G., Penke, L., & Wood, D. (2013). Self-regulation underlies temperament and personality: An integrative developmental framework. *Child Development Perspectives*, *7*, 255–260. https://doi.org/10.1111/cdep.12050

Denissen, J. J. A., Luhmann, M., Chung, J. M., & Bleidorn, W. (2019). Transactions between life events and personality traits across the adult lifespan. *Journal of Personality and Social Psychology, 116*, 612-633. http://doi.org/10.1037/pspp0000196

DeYoung, C. G. (2015). Cybernetic Big Five Theory. *Journal of Research in Personality*, *56*, 33–58. http://dx.doi.org/10.1016/j.jrp.2014.07.004

Donnellan, M. B., & Lucas, R. E. (2008). Age differences in the big five across the life span: Evidence from two national samples. *Psychology and Aging, 23*, 558–566. https://doi.org/10.1037/a0012897

Friedman, J., Hastie, T., Simon, N., & Tibshirani, R. (2016). glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models (Version 2.0-5). Retrieved from https://cran.r-project.org/web/packages/glmnet/index.html

Funder, D. C. (1991). Global Traits: A Neo-Allportian Approach to Personality. *Psychological Science*, *2*, 31–39. https://doi.org/10.1111/j.1467-9280.1991.tb00093.x

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. J. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe* (Vol. 7, pp. 7–28). Tilburg: Tilburg University Press.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40,* 84–96. https://doi.org/10.1016/j.jrp.2005.08.007

Goldberg, L. R. (2018). NEO-PI-R. *Harvard Dataverse*, V1, UNF:6:Sh2iB6vpAPrVxkaoewIqrg== [fileUNF]. https://doi.org/10.7910/DVN/HE6LJR

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12,* 55-67. https://doi.org/10.1080/00401706.1970.10488634

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Jackson, J. J., Walton, K. E., Harms, P. D., Bogg, T., Wood, D., Lodi-Smith, J., … Roberts, B. W. (2009). Not all Conscientiousness Scales Change Alike: A Multimethod, Multisample Study of Age Differences in the Facets of Conscientiousness. *Journal of Personality and Social Psychology, 96,* 446–459. https://doi.org/10.1037/a0014156

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, *51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003

Kosinski, M., Matz, S., Gosling, S., Popov, V. & Stillwell, D. (2015). Facebook as a Social Science Research Tool: Opportunities, Challenges, Ethical Considerations and Practical Guidelines. *American Psychologist, 70,* 543-556. https://doi.org/10.1037/a0039210

Lo, M.-T., Hinds, D. A., Tung, J. Y., Franz, C., Fan, C.-C., Wang, Y., … Chen, C.-H. (2017). Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nature Genetics*, *49*, 152–156. https://doi.org/10.1038/ng.3736

Lucas, R. E., & Donnellan, M. B. (2009). Age differences in personality: evidence from a nationally representative Australian sample. *Developmental Psychology, 45,* 1353–1363. https://doi.org/10.1037/a0013914

McAdams, D. P., & Pals, J. L. (2006). A new Big Five: Fundamental principles for an integrative science of personality. *American Psychologist*, *61*, 204–217. https://doi.org/10.1037/0003-066X.61.3.204

McCrae, R. R. (2015). A More Nuanced View of Reliability: Specificity in the Trait Hierarchy. *Personality and Social Psychology Review*, *19*, 97–112. https://doi.org/10.1177/1088868314541857

McCrae, R. R., & Costa, P. T. (2008). Empirical and theoretical status of the five-factor model of personality traits. In B. Boyle, G. Matthews, & D. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment: Volume 1 — Personality theories and models* (pp. 273–295). London: SAGE.

McCrae, R. R., & Costa, P. T. (2010). *NEO Inventories professional manual*. Odessa, FL: Psychological Assessment Resources.

McCrae, R. R., Costa, P. T., de Lima, M. P., Simões, A., Ostendorf, F., Angleitner, A., … Piedmont, R. L. (1999). Age differences in personality across the adult life span: Parallels in five cultures. *Developmental Psychology*, *35*, 466–477. https://doi.org/10.1037/0012-1649.35.2.466

McCrae, R. R., & Mõttus, R. (in press). A New Psychometrics: What Personality Scales Measure, with Implications For Theory and Assessment. *Current Directions in Psychological Science*.

McCrae, R. R., & Sutin, A. R. (2018). A Five-Factor Theory perspective on causal analysis. *European Journal of Personality*, *32*, 151-166. https://doi.org/10.1002/per.2134

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, *112*, 474–490. https://doi.org/10.1037/pspp0000100

Mõttus, R., Realo, A., Allik, J., Esko, T., Metspalu, A., & Johnson, W. (2015). Within-trait heterogeneity in age group differences in personality domains and facets: Implications for the development and coherence of personality traits. *PLoS ONE*, *10*, e0119667. https://doi.org/10.1371/journal.pone.0119667

Mõttus, R., Sinick, J., Terracciano, A., Hrebickova, M., Kandler, C., Ando, J., … Jang, K. (2018). Personality characteristics below facets: A replication and meta-analysis of cross-rater agreement, rank-order stability, heritability and utility of personality nuances. *Journal of Personality and Social Psychology*.

Plomin, R., & Stumm, S. von. (2018). The new genetics of intelligence. *Nature Reviews Genetics, 19,* 148-159. https://doi.org/10.1038/nrg.2017.104

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Revelle, W. (2018). psych: Procedures for Personality and Psychological Research (Version 1.8.12). Retrieved from http://CRAN.R-project.org/package=psych

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological Bulletin*, *126*, 3–25. http://doi.org/10.1037/0033-2909.126.1.3

Roberts, B. W., & Nickel, L. B. (2017). A critical evaluation of the Neo-Socioanalytic Model of personality. In J. Specht (Ed.), *Personality Development Across the Lifespan* (pp. 157–177). https://doi.org/10.1016/B978-0-12-804674-6.00011-9

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*, 1–25. https://doi.org/10.1037/0033-2909.132.1.1

Seeboth, A., & Mõttus, R. (2018). Successful Explanations Start with Accurate Descriptions: Questionnaire Items as Personality Markers for More Accurate Predictions. *European Journal of Personality*, *32*, 186–201. https://doi.org/10.1002/per.2147

Soto, C. J., & John, O. P. (2012). Development of Big-Five Domains and Facets in Adulthood: Mean-Level Age Trends and Broadly Versus Narrowly Acting Mechanisms. *Journal of Personality*, *80*, 881–914. https://doi.org/10.1111/j.1467-6494.2011.00752.x

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*, *100*, 330–348. https://doi.org/10.1037/a0021717

Terracciano, A., Costa, P. T., & McCrae, R. R. (2006). Personality Plasticity After Age 30. *Personality & Social Psychology Bulletin*, *32*, 999–1009. https://doi.org/10.1177/0146167206288599

Terracciano, A., McCrae, R. R., Brant, L. J., & Costa, P. T., Jr. (2005). Hierarchical linear modeling analyses of the NEO-PI-R scales in the Baltimore Longitudinal Study of Aging. *Psychology and Aging*, *20*, 493–506. https://doi.org/10.1037/0882-7974.20.3.493

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society: Series B, 73,* 273–282. https://doi.org/10.1111/j.1467-9868.2011.00771.x

Vainik, U., Dagher, A., Realo, A., Colodro-Conde, L., Mortensen, E. L., Jang, K., … Mõttus, R. (2019). Personality-obesity associations are driven by narrow traits: A meta-analysis. *Obesity Reviews, 20,* 1121–1131. https://doi.org/10.1111/obr.12856

Wagner, J., Lüdtke, O., & Robitzsch, A. (2019). Does personality become more stable with age? Disentangling state and trait effects for the big five across the life span using local structural equation modeling. *Journal of Personality and Social Psychology, 116,* 666-680. http://dx.doi.org/10.1037/pspp0000203

Waldmann, P., Mészáros, G., Gredler, B., Fuerst, C., & Sölkner, J. (2013). Evaluation of the lasso and the elastic net in genome-wide association studies. *Frontiers in Genetics, 4,* 270. https://doi.org/DOI: org/10.3389/fgene.2013.00270

Wilt, J., & Revelle, W. (2015). Affect, Behaviour, Cognition and Desire in the Big Five: An Analysis of Item Content and Structure. *European Journal of Personality, 29,* 478–497. https://doi.org/10.1002/per.2002

Wood, D., Nye, C. D., & Saucier, G. (2010). Identification and measurement of a more comprehensive set of person-descriptive trait markers from the English lexicon. *Journal of Research in Personality, 44,* 258–272. https://doi.org/10.1016/j.jrp.2010.02.003

Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science, 12,* 1100–1122. https://doi.org/10.1177/1745691617693393

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*, 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x